



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 1998

Qualität und Effizienz von Evaluationen: wie können Auswirkungen der Weiterbildung auf die Gesellschaft evaluiert werden?

Widmer, Thomas

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-159844>

Book Section

Published Version

Originally published at:

Widmer, Thomas (1998). Qualität und Effizienz von Evaluationen: wie können Auswirkungen der Weiterbildung auf die Gesellschaft evaluiert werden? In: Autonome Provinz Bozen-Südtirol; Abteilung deutsche und ladinische Schule und Kultur; Autonome Provinz Bozen-Südtirol; Amt für Weiterbildung; Deutsches Institut für Erwachsenenbildung. Qualitätsentwicklung durch Evaluation: Konzepte und Instrumente für die Weiterbildungspraxis. Bozen: s.n., 85-94.

Qualität und Effizienz von Evaluation: Wie können Auswirkungen der Weiterbildung auf die Gesellschaft evaluiert werden?

Dr. Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

Sehr geehrte Damen und Herren!

Es ist mir eine große Ehre, heute zum Thema "Qualität und Effizienz von Evaluation" sprechen zu dürfen. Ich habe die Einladung, an der 22. Tagung der Weiterbildung ein Referat zu halten, sehr gerne angenommen.

Ich möchte meine Präsentation folgendermaßen gestalten: In einem ersten Teil werde ich über zwei Forschungsprojekte berichten. Diese beiden Untersuchungen befassen sich mit der Qualität und der Effizienz von Evaluationen. Im zweiten Teil werde ich auf die Frage eingehen, wie die Auswirkungen der Weiterbildung auf die Gesellschaft evaluiert werden können. Dies ist eine Thematik, deren Behandlung für mich, einem Politikwissenschaftler, vor Ihnen, einem in Weiterbildungsfragen äußerst kompetenten Publikum, eine spezielle Herausforderung darstellt. Der zweite Teil meiner Ausführungen stützt sich neben den bereits erwähnten Forschungsprojekten auf Erfahrungen ab, die ich im Rahmen meiner Tätigkeit als Evaluator, Gutachter und Evaluationsberater machen konnte.

Qualität und Effizienz von Evaluation

Zuerst komme ich also zu den zwei Forschungsprojekten, die wir im Rahmen des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Maßnahmen" (NFP 27) durchgeführt haben und die durch den Schweizerischen Nationalfonds (SNF) gefördert wurden (siehe Widmer 1996a&b und Widmer/Rothmayr/Serdült 1996). In diesen beiden Untersuchungen wurden insgesamt fünfzehn Evaluationsstudien aus diversen Politikfeldern im Hinblick auf ihre Qualität und ihre Effizienz überprüft. Derartige Evaluationen von Evaluationsstudien werden auch als Meta-Evaluationen bezeichnet. Diese Beschäftigung mag Ihnen reichlich akademisch erscheinen, doch ich werde im folgenden zu zeigen versuchen, daß sich daraus durchaus praxisrelevante und hoffentlich auch handlungsrelevante Erkenntnisse gewinnen lassen.

Die Qualitätskriterien

Die Evaluationsforschung hat in anderen Ländern schon früh erkannt, daß sie in der ideologischen Steuerungsdebatte und in der aktuellen Diskussion zwischen finanzpolitisch orientierten Sporbemühungen und sozialstaatlichen Sicherungsanstrengungen instrumentalisiert werden kann. Um die Gefahr von Alibi-Evaluation zu bannen, haben deshalb erfahrene amerikanische Evaluatoren schon im Jahr 1981 Standards entwickelt, welche die Anforderungen an gute Evaluation systematisieren. Diese Kriterien sind in jüngster Zeit überarbeitet worden (Joint Committee 1994). Wir paßten diese Standards an die spezifisch schweizerischen Verhältnisse an und setzten die insgesamt dreißig Einzelkriterien zur Qualitätsbewertung der fünfzehn Evaluationsstudien ein. Sie erlauben es, Evaluationen aus vier Dimensionen zu beurteilen:

- Die erste Dimension betrifft die *Nützlichkeit* von Evaluationen. Die entsprechenden Standards sollen dafür sorgen, daß Evaluationen die Informationsbedürfnisse aller Beteiligten auf Seiten von Politik und Verwaltung wie auf Seiten der Adressaten und Programmbetroffenen befriedigen. Diese Standards verlangen hohe Kompetenz und Glaubwürdigkeit der Evaluatoren und setzen voraus, daß Evaluationsberichte umfassend und rechtzeitig alle Interessierten orientieren.
- Mit der zweiten Dimension wird die *Anwendbarkeit* angesprochen. Evaluationen sollen praktikabel und politisch tragbar sein. Dazu gehört auch, daß der Wert der gewonnenen Informationen die Kosten rechtfertigt.

- Die dritte Dimension zielt auf *Korrektheit*, d.h. auf die ethische und rechtliche Verantwortlichkeit in der Evaluationsforschung. Wechselseitige Verpflichtungen sind detailliert in einem Vertrag festzuhalten und einzuhalten. Menschenrechte und Menschenwürde sind zu respektieren. Evaluationsberichte sollen offen, direkt, ehrlich und fair sein und sowohl auf Stärken wie auf Schwächen des Evaluationsgegenstandes hinweisen.
- Als vierte Dimension gilt es, das Gebot der *Genauigkeit* zu erwähnen. Die entsprechenden Standards beziehen sich direkt auf die Einhaltung bestimmter Regeln der sozialwissenschaftlichen Arbeitsweise. Dabei sind Standards für quantitative *und* für qualitative Vorgehensweisen, für empirisch-analytische *und* für alternative Forschungsansätze ausgearbeitet worden. Genaues und regelkonformes Arbeiten sowie eine detaillierte methodische Berichterstattung werden für alle Arten von Evaluationsforschung gefordert.

Vorsorglich ist darauf hinzuweisen, daß in der Regel nicht allen Standards vollumfänglich Genüge getan werden kann. Vielmehr bestehen zwischen verschiedenen Standards Widersprüche. So kann eine umfassende Informationsbeschaffung allzu hohe Kosten verursachen. So kann eine vollständige Information der Öffentlichkeit die Menschenwürde eines einzelnen verletzen. So kann der Einsatz von aufwendigen Verfahren zur Erhöhung der Zuverlässigkeit von Interpretationen zu einem verspäteten Abschluß der Arbeiten führen. In solchen Fällen sind sorgfältige Abwägungen vorzunehmen, und es sind höchstens Optimierungen der Evaluationen zu erreichen.

Das hier vorgestellte Kriterienraster eignet sich nicht nur für die (nachträgliche) Bewertung von Evaluationsstudien in einer Meta-Evaluation, sondern kann auch als Orientierungslinie für die praktische Durchführung von Evaluationen dienen oder als Hilfsmittel bei der Aus- und Weiterbildung in Evaluationsforschung eingesetzt werden.

Die untersuchten Evaluationsstudien

Die in der Qualitätsprüfung betrachteten fünfzehn Evaluationsstudien decken ein sehr breites Spektrum an Themenbereichen ab, wobei der Bildungsbereich und die Gesundheitspolitik explizit ausgeschlossen wurden. Die Evaluationen sind ausnahmslos externe Evaluationen, das heißt: Die Evaluationen sind als Aufträge vergeben worden. Im einzelnen sind die folgenden Studien untersucht worden:

Langevaluationen:

- Die erste Studie trägt den Titel "Umweltpolitik und technische Entwicklung". Es handelt sich dabei um eine Evaluation, welche die Auswirkungen umweltpolitischer Vorschriften für Hausfeuerungen auf Innovationen im Bereich der Heizungstechnik untersucht. (1)
- Die Evaluation "Unternehmerische Innovationsprozesse" fragt nach den innovationsfördernden Auswirkungen regionaler Wirtschaftsförderungsmaßnahmen. (2)
- Die Studie "Bedingt wirksam: Das Lohnungleichheitspostulat des Heimarbeitsgesetzes" untersucht eine Gesetzesvorschrift zum Schutz der Arbeitnehmer und vor allem der Arbeitnehmerinnen, die im Bereich der Heimarbeit tätig sind. (3)
- Die vierte Studie befaßt sich mit den Auswirkungen der Wohneigentumsförderung durch die schweizerische Eidgenossenschaft, ein Programm, das zur Erhöhung der in der Schweiz traditionell sehr tiefen Wohneigentumsquote beitragen soll. (4)
- Die nächste Studie untersucht, ob die Einführung von Gebühren für die Kehrrichtentsorgung in Schweizer Gemeinden zu einer Reduktion der Kehrrichtmenge geführt hat. (7)
- Mit einer ähnlichen Frage befaßt sich auch die Studie "Verbrauchsabhängige Heizkostenabrechnung - Förderung erneuerbarer Energien", wobei hier die Auswirkungen einer Vorschrift untersucht werden, welche die individuellen Heizkosten nach dem Verursacherprinzip den einzelnen Personen anlastet. (8)
- Bei der Studie "Evaluation des BUWAL" handelt es sich um eine Organisationsanalyse, welche die Tätigkeiten des schweizerischen Bundesamtes für Umwelt, Wald und Landschaft untersucht, also um eine Studie, die sich mit einer Organisation und nicht mit politischen Programmen oder Maßnahmen befaßt. (9)

Kurzevaluationen:

- Eine weitere der untersuchten Evaluationen führt eine Effizienzanalyse der Züricher Gemeinschaftszentren durch. Bei den Züricher Gemeinschaftszentren handelt es sich um Institutionen eines privaten Trägers, die in der quartierbezogenen Gemeinwesenarbeit tätig sind. Auch diese Studie ist also eine institutionelle Evaluation. (5)
- Die Evaluation "Schneeräumung auf Urner Paßstraßen" sucht nach effizienteren Lösungen für die Schneeräumung im Bergkanton Uri. (6)
- In der Studie "Evaluierung der EWR-Informationskampagne" geht es um die Auswirkungen der Informationskampagne, welche die Bundesregierung im Vorfeld der Volksabstimmung über einen Beitritt der Schweiz zum Europäischen Wirtschaftsraum lanciert hatte. (10)
- Eine weitere Evaluation analysiert die Auswirkungen eines Investitionshilfeprogramms für den landwirtschaftlichen Hochbau. Dabei interessieren in erster Linie die Auswirkungen auf den Strukturwandel in der Landwirtschaft. (11)
- Die nächste Studie befaßt sich mit dem Submissionswesen des Kantons Bern, also mit der Vergabe öffentlicher Aufträge in diesem Kanton. (12)
- Die "Effizienz der Maßnahme zur Abwehr von Naturgefahren im Kanton Wallis", gemeint sind damit die Schutzverbauungen gegen Lawinen, Murgänge usw., steht im Zentrum einer weiteren Studie. (13)
- Verschiedene Maßnahmen der Wirtschaftsförderungen des französischsprachigen Kantons Neuchâtel wurden in einer der Studien einer Evaluation unterzogen. (14)
- Die letzte der evaluierten Evaluationen befaßt sich mit dem "Design Preis Schweiz", wobei vor allem die wirtschaftlichen und kulturellen Auswirkungen dieses Design-Wettbewerbs genauer betrachtet wurden. (15)

Wie aus dieser Aufzählung hervorgeht, sind die Evaluationsstudien in ganz unterschiedlichen Kontexten entstanden. Die Komplexität der Evaluationsgegenstände variiert von Studie zu Studie deutlich. Die Tragweite der untersuchten Fragestellungen ist sehr unterschiedlich. Aber auch der Umfang der Evaluationsprojekte, also primär die finanziellen und zeitlichen Ressourcen, ist sehr verschieden.

Die Abbildung 1 zeigt die Verteilung der Studien hinsichtlich ihrer Kosten und ihres Zeitbedarfs. Um einem Vergleich zwischen den in finanzieller und zeitlicher Hinsicht aufwendigen und weniger aufwendigen Studien zu ermöglichen, haben wir die fünfzehn Evaluationsstudien in zwei Gruppen aufgeteilt.

Die Unterscheidung in sogenannte Kurzevaluationen und in die umfangreicheren Langevaluationen richtet sich nach dem Zeit- und dem Kostenbedarf der Studien. Die Grenze, die wir gezogen haben, liegt bei einem Jahr und bei 100.000,- Schweizer Franken. Dieser Frankenbetrag entspricht etwa 115 Millionen Lire, wobei ich bei der Umrechnung die Kaufkraftunterschiede *nicht* berücksichtigt habe. Der Übergang von den kleineren zu den größeren Studien verläuft fließend. Die Schwelle für die Zuordnung zu den beiden Gruppen haben wir pragmatisch gewählt. Weiter ist zu betonen, daß auch günstigere Studien länger als ein Jahr dauern oder daß kurzfristige Studien teuer sein können.

Ergebnisse

Grundsätzlich haben unsere Abklärungen gezeigt, daß die Qualität der untersuchten Evaluationsstudien als gut bezeichnet werden kann. Trotz des generell positiven Befundes, sind wir im Rahmen der beiden Meta-Evaluationsstudien auch auf gewisse Schwachpunkte gestoßen. Diese Tabelle (vgl. Abbildung 2) zeigt, in welchen Bereichen die Evaluationen auffällig gut bewertet wurden und wo wir häufiger auf Schwachstellen gestoßen sind. Betrachtet man die Zugehörigkeit der aufgeführten Kriterien, wird deutlich, daß die Bereiche Nützlichkeit, Anwendbarkeit und Korrektheit positiv bewertet wurden, während sich unter den eher negativ bewerteten Kriterien ausnahmslos solche aus der Gruppe 4 zur Genauigkeit befinden. Die Meta-Evaluation hat relativ viele Verstöße gegen die Genauigkeitsstandards ausgemacht und damit selbstverständlich auch da und dort in ein Wespennest gestochen. Dazu möchte ich folgendermaßen Stellung nehmen:

- In einzelnen Fällen mußten eigentliche Mängel bei der Umsetzung eines Designs oder bei der Anwendung einer bestimmten Methode gerügt werden. Dazu ist zu sagen, daß Fehler gerade in einem für viele neuen Gebiet wie der Evaluation noch leichter entstehen können als sonst. Gegenseitige konstruktive Kritik muß hier zur wissenschaftlichen Kultur gehören und kann alle weiterbringen.
- Das Hauptproblem der Evaluationsforschung in der Schweiz besteht indessen nach unseren Beobachtungen darin, daß verschiedene Equipen sich ausschließlich einer einzigen Methode verschreiben. Je nach Evaluationsfrage ist aber die eine oder die andere Methode besser geeignet, schlüssige Antworten zu finden. Nicht selten drängt es sich auf, einen Methodenmix einzusetzen. Dies ist indessen nur möglich, wenn Evaluationsteams bereit und in der Lage sind, pragmatisch die geeigneten Methoden aus einem reichhaltigen Repertoire zu holen und technisch korrekt einzusetzen. Diesbezüglich hat die schweizerische Evaluationsforschung noch einen Schritt nach vorne zu machen.

Der Vergleich zwischen den umfangreicheren sogenannten Langevaluationen und den kleineren Kurzevaluationen erlaubte uns die Definition von Bedingungen, unter denen das Risiko, daß Kurzevaluationen unbefriedigende Resultate erbringen, klein gehalten werden kann. Diese Erfolgsbedingungen für Kurzevaluationen lassen sich wie folgt umschreiben:

Hinsichtlich des Evaluandums:

- Evaluanda ohne allzu große politische, soziale, ethische und finanzielle Relevanz
- Evaluanda, zu denen bereits ausreichend Vorkenntnisse vorliegen
- Evaluanda mit eher geringer Komplexität
- Evaluanda, deren Ausgestaltung/Umsetzung unter Berücksichtigung evaluationsspezifischer Erfordernisse erfolgt
- Evaluanda, bei denen ein hoher Zeitdruck besteht

Hinsichtlich der Evaluation:

- Evaluationen mit hoher Akzeptanz
- Evaluationen mit präziser Planung
- Evaluationen mit formativem, nicht mit summativem Ansatz
- Evaluationen, die sich mit Prozessen und weniger mit 'impacts' befassen
- Evaluationen, von denen keine Vertiefung und keine Innovationen erwartet werden

Hinsichtlich der Durchführung der Evaluation:

- enge Projektbegleitung
- einfache Projektorganisation
- außerordentlich kompetente und flexible Evaluierende

Hinsichtlich der Konsequenzen der Evaluation:

- hohe Risikobereitschaft der Beteiligten und Betroffenen

Sind diese Bedingungen gegeben, können Kurzevaluationen eine Alternative zu umfangreicheren Evaluationen darstellen, indem sie in kurzer Zeit und mit geringen Kosten zu Ergebnissen führen, die methodisch ausreichend abgesichert sind. Sind die angeführten Bedingungen nicht erfüllt, sind umfangreichere Evaluationsstudien vorzuziehen, weil die mit Kurzevaluationen verbundenen Risiken zu hoch sind.

Zur Evaluation der gesellschaftlichen Auswirkungen der Weiterbildung

Spezifität des Bildungsbereichs

Personen aus dem Bildungsbereich weisen immer wieder darauf hin, daß sich Modelle aus anderen Bereichen nicht auf die Bildung übertragen lassen. Dafür werden verschiedene Argumente angeführt, die auf die spezifische Qualität des Bildungsbereiches verweisen, wie etwa die hohe Komplexität der Prozesse und die herausragende Bedeutung sozialer Interaktionen. Ich möchte an dieser Stelle keinesfalls eine Gegenposition vertreten. Ich möchte lediglich darauf hinweisen, daß diese Situation nicht alleine im Bereich der Bildung besteht. So treffe ich analoge Argumente auch in anderen Sektoren, etwa im Bereich der Kinder- und

Jugendhilfe, der öffentlichen Fürsorge, der Sozialarbeit, aber auch in der Wissenschaft und im Gesundheitswesen. Beratende, problemlösende, zwischenmenschliche, kognitive und edukative Prozesse scheinen Zugriffe technokratischer Art nicht zu dulden. Diesen - meines Erachtens durchaus verständlichen - Widerstand führe ich primär auf die Angst zurück, daß einem fremdbestimmte, oft wenig adäquate sozialtechnologische Modelle von außen aufgezwungen werden. Diese Befürchtungen verschärfen sich zusätzlich, wenn die Situation von finanziellen Engpässen und Sparanstrengungen begleitet wird. Es wird befürchtet, daß die eigene Fachkompetenz als Lehrperson, als Jugendhelfer, als Wissenschaftler oder Psychologin - zumeist ein genuiner Bestandteil der professionellen Selbstidentität - von betriebswirtschaftlichen Kostenrechnungen und technischen Leistungsindikatoren überdeckt oder sogar verdrängt wird (vgl. dazu die Beiträge in Heiner 1996). Nur schon aus diesem einen Grund ist es für die Evaluation unabdingbar, sich mit den relevanten Werten, Einstellungen und Interessen vertieft auseinanderzusetzen und Evaluationskonzepte zu entwickeln, die den spezifischen Anforderungen gerecht werden.

Wirkungsevaluationen

Im Gegensatz zu Prozeßevaluationen, welche die Prozesse einer Bewertung unterziehen, konzentriert sich die Wirkungsevaluation auf die Auswirkungen des zu evaluierenden Objekts, das wir auch als Evaluandum bezeichnen. Dabei versucht sie aufgrund von sogenannten Wirkungsketten die kausalen Zusammenhänge zwischen Evaluandum und den Auswirkungen zu erfassen.

Schematisch lassen sich dabei folgende Stufen der Wirkungskette unterscheiden. Ich verwende hier absichtlich die englischen Begriffe, um Mißverständnisse zu vermeiden. Das Verständnis sowie die Verwendung der deutschen Begriffe ist hier uneinheitlich und wenig gefestigt.

Wie aus dieser Zusammenstellung hervorgeht, stellt die sogenannte "impact"-Evaluation äußerst hohe Ansprüche. Der wissenschaftliche Nachweis, daß eine gesellschaftliche Veränderung ursächlich auf eine bestimmte Tätigkeit, ein bestimmtes Programm oder eine bestimmte Institution zurückgeführt werden kann, ist nur sehr schwer zu erbringen. Ich möchte dies anhand eines simulierten Evaluationsszenarios erläutern: Die Teilnehmerinnen und Teilnehmer eines Englischkurses werden zehn Jahre nach der Absolvierung des Kurses daraufhin befragt, wie häufig sie heute Gelegenheit haben, die englische Sprache einzusetzen. Eine derartige Untersuchungsanlage ist mit verschiedenen Schwächen verbunden, so zum Beispiel:

- Viele der Teilnehmerinnen und Teilnehmer können aus verschiedenen Gründen durch die Befragung nicht mehr erreicht werden.
- Der Spracheinsatz vieler Personen wird maßgeblich von Faktoren beeinflusst, die in keiner Weise einen Zusammenhang mit dem interessierenden Sprachkurs aufweisen. So wäre es vermessen, den Einsatz der englischen Sprache einer Person, die vor sieben Jahren nach Australien ausgewandert ist, auf den früher besuchten Sprachkurs zurückzuführen. Zudem haben möglicherweise verschiedene der Teilnehmerinnen und Teilnehmer in den letzten zehn Jahren weitere Englischkurse belegt. Dies erschwert die kausale Zuordnung der Wirkungen zu einem bestimmten Sprachkurs erheblich.
- Die Resultate dieser Studie werden kaum von großem Interesse sein, verändert sich doch die Ausgestaltung der Bildungsaktivitäten laufend. Oder führt eine Ihrer Institutionen seit zehn Jahren denselben Kurs in exakt identischer Form mit der gleichen personellen Besetzung durch? Die Untersuchung wird also Ergebnisse hervorbringen, die für die heutigen Aktivitäten ohne Bedeutung sind. Bei Bildungsprogrammen handelt es sich für die Evaluation um - verzeihen Sie mir die militärische Ausdrucksweise - "moving targets".
- Um Aussagen darüber zu machen, ob der erwähnte Sprachkurs etwas bewirkt hat, müßte eine Vergleichsebene bestehen. Es wird aber äußerst schwierig sein, eine Gruppe von Personen zu bestimmen, die sich - bis auf die Nichtteilnahme am Sprachkurs - in keiner Weise von den Teilnehmerinnen und Teilnehmern unterscheiden und somit eine verlässliche Kontrollgruppe darstellen würden.

Man wird einwenden, daß viele der angeführten Schwierigkeiten dadurch bereinigt werden können, daß weniger als zehn Jahre zugewartet wird, um die Befragung durchzuführen. Dieser Einwand ist jedoch nicht stichhaltig, denn Bildungsaktivitäten zeichnen sich ja gerade durch ihre längerfristige Wirkungsstruktur aus. Dies wird durch die Erfordernisse des lebenslangen Lernens noch akzentuiert. Gefragt sind heute nachhaltige Wissensbestände mit langer Lebensdauer: dem Strukturwissen kommt heute größere Bedeutung zu als dem Faktenwissen (Dubs 1995: 166).

Das Hauptproblem in der skizzierten Untersuchungsanlage besteht aber darin, daß damit gar nicht die gesellschaftlichen Auswirkungen des Sprachkurses erfaßt werden können. Diese Untersuchung zielt darauf ab, den längerfristigen "outcome" des Sprachkurses zu ermitteln und nicht dessen "impact".

Versucht man eine Evaluation zu konzipieren, die auf die Erfassung der gesamtgesellschaftlichen Auswirkungen des Sprachkurses ausgerichtet ist, wird man mit weiteren Schwierigkeiten konfrontiert. Nur schon die Kleinheit der Gruppe der Teilnehmerinnen und Teilnehmer - zwanzig oder dreißig Personen - in Relation zur Gesamtbevölkerung wird es unmöglich machen, die gesamtgesellschaftlichen "impacts" eines einzelnen Kurses zu erfassen. Eher denkbar wäre es, die "impacts" des gesamten Weiterbildungsangebotes auf gesamtgesellschaftliche Auswirkungen zu untersuchen. Dazu muß jedoch gesagt werden, daß durch die Resultate einer in dieser Art konzipierten Evaluation kaum ein Beitrag zur Qualitätsentwicklung geleistet werden kann. Die zu erwartenden Ergebnisse sind zu pauschal, um daraus Hinweise für die Qualitätssicherung oder die Qualitätsverbesserung der einzelnen Weiterbildungsangebote zu erhalten. Eine solche Untersuchung könnte aber wichtige Informationen zur Gesamtkonzeption der Weiterbildung bereitstellen. Speziell denke ich dabei an potentielle Angebotslücken, an Überangebote in bestimmten Bereichen oder an nicht zielkonforme Nebenwirkungen.

Aus den dargelegten Überlegungen schließe ich, daß es, bezogen auf die einzelne Weiterbildungseinrichtung oder das einzelne Weiterbildungsangebot, aussichtsreicher ist, Evaluationen auf der Ebene der Prozesse, der "outputs" und eventuell der "outcomes" zu situieren. "Impact"-Evaluationen sind eher in Globalevaluationen anzustreben, wobei in diesem Fall umfangreiche Evaluationen zu planen sind, um haltbare Resultate erarbeiten zu können.

Pluralismus und Veränderungen der Interessen

An die Weiterbildung werden von unterschiedlicher Seite Interessen, Anforderungen, Wünsche und Erwartungen gerichtet. Erwähnen möchte ich hier zum einen die direkt beteiligten Kreise, also etwa die Teilnehmerinnen und Teilnehmer, die Mitarbeiterinnen und Mitarbeiter sowie Leitungen der in der Weiterbildung tätigen Institutionen und die finanzierenden öffentlichen oder privaten Stellen. Nicht zu vergessen sind aber auch die durch die Weiterbildungstätigkeit Betroffenen, also das soziale Umfeld der oben erwähnten Kreise bis hin zur gesamten Bevölkerung. Die Beteiligten und Betroffenen, also alle, die gewisse Interessen am Angebot haben, werden in der Fachsprache auch als "stakeholders" bezeichnet.

Alle diese Personen und Institutionen haben ihre spezifische Auffassung, was eine gute Weiterbildung ausmacht. Diese Interessen können im Einzelfall auch übereinstimmen; es ist aber damit zu rechnen, daß dies häufig nicht der Fall sein wird. Vielmehr wird man zumeist auf Interessen stoßen, die sich widersprechen. Weiter ist davon auszugehen, daß die in irgendeiner Weise direkt am Weiterbildungsangebot Beteiligten ihre eigenen Interessen eher einbringen können, wobei von einem teilweise markanten Machtgefälle auszugehen ist. Die ohne direkten Zugang Betroffenen haben dagegen zumeist keine Möglichkeit, Einfluß zu nehmen.

Eine aussagekräftige Evaluation hat die unterschiedlichen mit der Weiterbildung verbundenen Interessen zu berücksichtigen. Sie sollte sich nicht darauf reduzieren, die Sicht eines bestimmten Personenkreises zu privilegieren. Insbesondere möchte ich davor warnen, die reine Innensicht zu stark zu betonen. Aus der Gegenüberstellung von Innen- und Außenperspektive ergeben sich oft die fruchtbarsten Erkenntnisse.

Ein Aspekt, der bei Evaluationen leider oft vernachlässigt wird, ist die Frage nach der Übereinstimmung zwischen der angestrebten Zielgruppe und der effektiv erreichten Adressatengruppe. Anders ausgedrückt: Gibt es in der Zielgruppe Nichtteilnehmer? Die Vernachlässigung dieses Aspekts kann dazu führen, daß das Wirkungspotential eines Weiterbildungsangebotes falsch eingeschätzt wird.

Die Interessen der Beteiligten und Betroffenen sind aber nicht nur sehr vielfältig, sie sind auch ständiger Veränderung ausgesetzt. So verändert sich die Haltung der finanzierenden Stellen in Zeiten knapper Ressourcen maßgeblich. Die Geldgeber werden in Zeiten knapper Ressourcen vermehrt beabsichtigen, die Kosten zu reduzieren, auch wenn dies mit Qualitätseinbußen verbunden sein könnte. Aber auch kurzfristig können sich die Einstellungen verändern. Beispielsweise wird ein Kursteilnehmer mit der Erwartungshaltung, seine Sprachkenntnisse zu verbessern, einen Sprachkurs antreten. Schon nach kurzer Zeit kann sich aber die Gewichtung der Wertedimensionen deutlich verschieben, so daß der Lernerfolg zugunsten sozialer Interaktion deutlich in den Hintergrund tritt. Dies wird sich nach Abschluß des Kurses möglicherweise wieder verändern, wenn wir von der Voraussetzung ausgehen, daß der Teilnehmer in diesem Kurs *nicht* seine Lebensgefährtin kennengelernt hat.

Konsequenzen

Als Konsequenz ergibt sich für die Evaluation erstens, daß die Bestimmung der Bewertungskriterien sehr große Aufmerksamkeit erfordert. Aufgrund divergierender Interessen verfolgen die Akteure unterschiedliche Zielsetzungen. Daraus können komplexe Zielsysteme resultieren, die durch die Evaluation aufgearbeitet werden müssen, um fundierte und sinnvolle Aussagen zu ermöglichen.

Zweitens ist davon auszugehen, daß es nicht ausreichen wird, nur gerade eine Evaluationsmethode einzusetzen. Zumeist wird man zweckmäßigerweise eine Kombination unterschiedlicher Ansätze, Methoden und Techniken favorisieren (Klöti/Widmer 1997, Widmer/Binder 1997, vgl. Moser/Wettstein 1997).

Zum Dritten wird man, wenn immer möglich, eine Kombination von Selbst- und Fremdevaluation ins Auge fassen. Diese Vorgehensweise vermindert die Gefahr, daß die Vielfalt von Werten und Zielen außer acht gelassen wird. Sie gestattet eine Kombination der Innen- und Außensicht, die wertvolle Hinweise und Anregungen erlaubt. Sie reduziert zudem auch die - zumindest in der Schweiz - in ähnlichen settings oft gehegte Befürchtungen, daß Evaluationen zu Kontrollzwecken eingesetzt werden. Vielmehr sollten Evaluationen aber einen Beitrag zur Qualitätsentwicklung leisten. Kontroll- und Aufsichtsfunktionen sind für diesen Zweck strikt von Evaluationsfunktionen zu trennen.

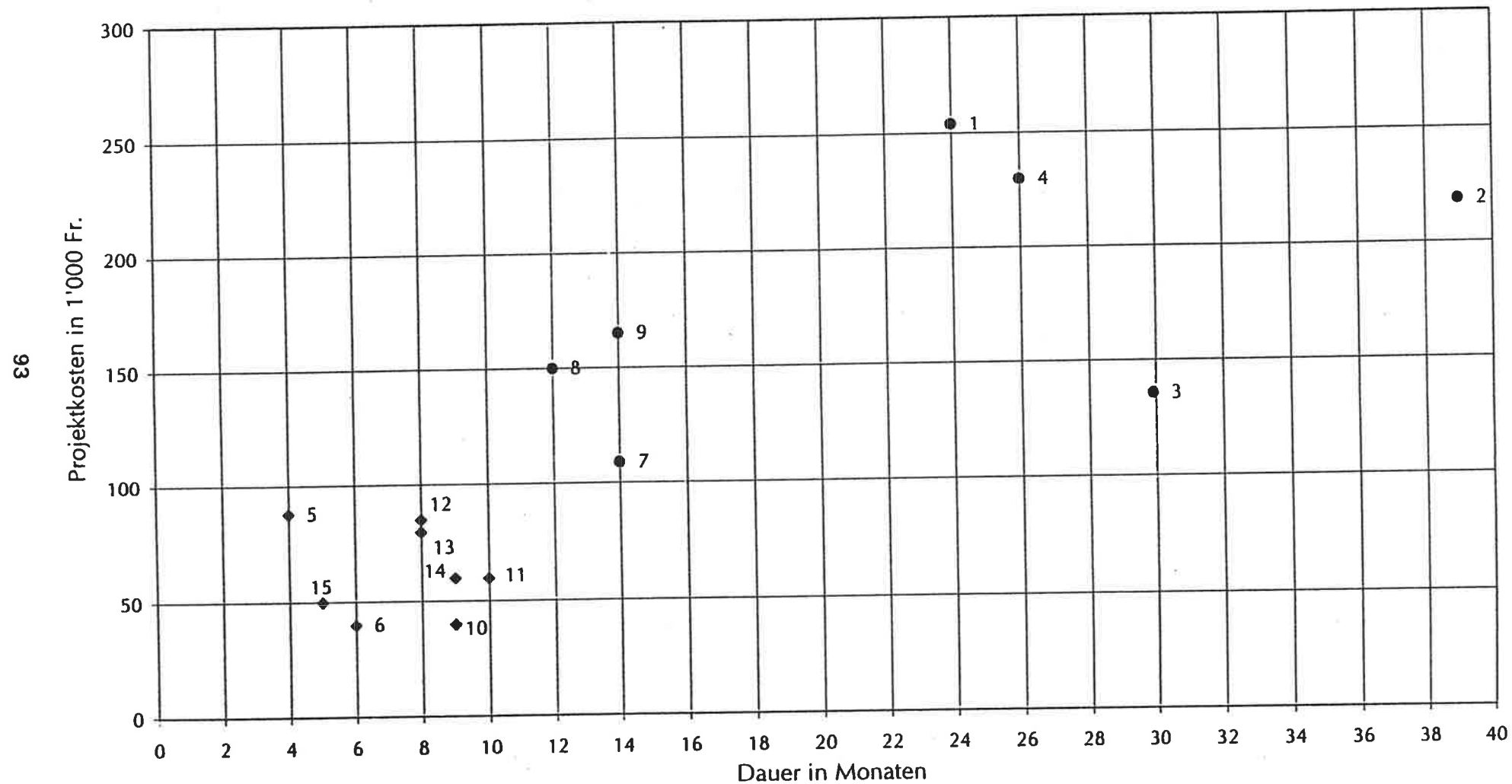
Um die Qualitätsentwicklung zu fördern - und das wäre der vierte Punkt -, sind Evaluationen in Zusammenarbeit mit den durch die Evaluation betroffenen Personen auszugestalten. Es ist ein sogenannter partizipativer Evaluationsansatz zu bevorzugen, der die Steuerung der Evaluation mehr oder weniger stark den Betroffenen überträgt.

Der fünfte - und letzte - Punkt betrifft die Terminierung von Evaluationen. Damit die Evaluation einen Beitrag zur kontinuierlichen Qualitätsentwicklung zu leisten vermag, sollte sie nicht als einmaliges Ereignis, sondern vielmehr als dauerhafter Prozeß aufgefaßt werden. Die Schulentwicklung kann so maßgeblich gefördert werden.

Literatur

- Dubs, Rolf (1995): Lehrerverhalten. Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht. Zürich: SKV.
- Heiner, Maja (Hrsg.) (1996): Qualitätsentwicklung durch Evaluation. Freiburg: Lambertus.
- Joint Committee on Standards for Educational Evaluation (1994): The Program Evaluation Standards. 2nd Edition. Thousand Oaks: Sage.
- Klöti, Ulrich/Widmer, Thomas (1997): Untersuchungsdesigns. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn.
- Moser, Heinz/Wettstein, Heinz (1997): Evaluation und Schulentwicklung - fünf Thesen. schweizer schule, 2/1997: 22-30.
- Widmer, Thomas (1996a): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Bern: Haupt.
- Widmer, Thomas (1996b): Fallstudien zur Meta-Evaluation. 7 Bde. Serie Schlußberichte. Nationales Forschungsprogramm "Wirksamkeit staatlicher Maßnahmen" (NFP 27). Bern: NFP 27.
- Widmer, Thomas/Binder, Hans-Martin (1997): Forschungsmethoden. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn.
- Widmer, Thomas/Rothmayr, Christine/Serdült, Uwe (1996): Kurz und gut? Zur Qualität und Effizienz von Kurzevaluationen. Chur: Rüegger.

Dauer und Kosten der fünfzehn Evaluationen



***Kriterien mit überdurchschnittlich positiver bzw.
negativer Bewertung***

<i>Positive Bewertung</i>	<i>Negative Bewertung</i>
1.2. Glaubwürdigkeit des Evaluators	4.3. Beschreibung von Zielen und Vorgehen
1.3. Informationsrahmen und -auswahl	4.4. Verlässliche Informationsquellen
1.6. Verbreitung des Berichts	4.5. Valide Messung
1.7. Rechtzeitigkeit des Berichts	4.6. Reliable Messung
2.1. Praktische Verfahren	4.8. Analyse quantitativer Informationen
3.1. Formale Verpflichtung	4.10. Gerechtfertigte Folgerungen
3.5. Schutz individueller Menschenrechte	
3.7. Ausgewogene Berichterstattung	